IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

I, SALLY LESLEY HEDLEY, B.A., M.I.L., M.I.T.I., declare

1. That I am a citizen of the United Kingdom of Great Britain and Northern Ireland, residing at 29 Parkholme Road, London E8 3AG, United Kingdom.

2. That I am well acquainted with the German and English languages.

3. That the attached is a true translation into the English language of the specification of International Patent Application No. PCT/EP00/00462.

4. That all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements are made with the knowledge that wilful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such wilful false statements may jeopardise the validity of the patent application in the United States of America or any patent issuing thereon.

Declared this 11th day of July, 2001.

Sally Hedley.

SALLY HEDLEY

Process for the combinatorial discovery of reactions for the preparation of useful products.

The present invention relates to the discovery and preparation of chemical compounds having desired and useful physical, chemical and/or biological properties by means of an iterative process based on multicomponent reactions. The compounds according to the invention can be used as medicaments, veterinary products, vaccines, cosmetics, plant protection preparations etc. or as additives thereto or as ligands, catalysts, catalytic cofactors, detector molecules, polymers, peptides and adhesives.

Processes for discovering new chemical compounds having desired physical, chemical and biological properties and new reactions for the preparation of chemical compounds having desired physical, chemical and biological properties are the subject of numerous patents, procedures, methods and scientific research projects. Such processes are intended especially to achieve one or more of the following objectives:

(a) the generation of new chemical reactions, basic structures, compounds or combinatorial substance libraries,

(b) the generation of a high degree of chemical diversity, the term "diversity" being defined as the information content of a chemical reaction, compound or substance library,

(c) the provision of processes for generating combinatorial substance libraries,

(d) the provision of optimisation and other processes for discovering active compounds from such libraries,

(e) the provision of new biological test systems and processes,

(f) the provision of processes for the synthesis of desired compounds,

(g) the analysis and optimisation of the duration of individual steps in the discovery and preparation of such compounds, and

(h) the analysis and optimisation of the costs of the individual steps in the discovery and preparation of such compounds.

None of the processes known hitherto, however, achieves all of the above-mentioned objectives (a-h) simultaneously. The goal of a fast and efficient way of discovering and preparing useful chemical compounds is therefore at best only partially achieved by the prior art:

The (targeted) synthesis of combinatorial substance libraries has been described as a route to discovering new chemical compounds having desired properties (Gallop, *Journal of Medicinal Chemistry*, 37(9), **1994**, 1233-1250). That method delivers a large number of new chemical compounds - a substance library - in which generally only the substituents around a common chemical basic scaffold are varied.

Furthermore, in that method the libraries are built up using a limited number of sequential reactions, which allows only a low degree of diversity of the basic scaffolds used. Numerous molecular properties, such as, for example, lipophilicity, oral bioavailability, biological activity, metabolic stability etc.,

are associated with those basic scaffolds, however, so that many of those properties cannot be obtained with those substance libraries.

The general diversity, that is to say the information content of those systematic substance libraries, is therefore low in comparison with random substance libraries on account of the principle by which they are constructed. (The term "random substance libraries" is here used to denote collections of substances that are not capable of being prepared by a uniform, systematic process, such as, for example, collections of natural substances.) Such combinatorial substance libraries therefore have, on account of their low degree of diversity and their redundant information content, the disadvantage that there is less probability of finding in them interesting chemical compounds having the desired biological activity. They also have disadvantages in respect of the time required and the costs incurred for preparing and testing the compounds.

The low degree of diversity also has advantages, however: since the compounds are always prepared and tested as chemically related families, the structure-activity relationships (SARs) obtained in each case are limited. Those SARs enable the errors frequently occurring in biological testing to be excluded, such as, for example, false positive or false negative signals which may arise as a result of apparatus defects, impurities in the test samples etc.. Furthermore, such an SAR can provide pointers to a potential optimisation of the chemical compounds in respect of their biological and other properties.

Random substance libraries generally do not exhibit the above-described disadvantages of a low degree of diversity (J.P. Devlin, The Discovery of Bioactive Substances - High Throughput

Screening, Marcel Dekkert, **1998**). The broad testing of such substance libraries for various biological activities is therefore a standard route to discovering biologically active compounds. A disadvantage of that process, however, is that the compounds so found often cannot be prepared by a simple, efficient and fast synthesis. The formulation of a combinatorial synthesis as a means of access to those compounds for the optimisation of the substance properties is time-consuming and expensive. A further disadvantage of such random substance libraries, from the standpoint of their biological testing, lies in the fact that they are not built up systematically, which does not enable false positive or false negative test results to be excluded on the basis of SARs. Such SARs have to be built up only in a subsequent step, after the testing of those libraries, by the synthesis of chemically related compounds. If the testing of random libraries has resulted in a large number of chemically diverse compounds having interesting biological activities, that subsequent step is extremely time-consuming because the preparation of each one of those compounds requires the use of other chemical processes which are not always known. That process therefore often results in the selection of compounds that are chemically easier to prepare, while more complicated compounds, such as, for example, natural substances, are not given any further consideration.

Processes for the discovery and preparation of biologically active compounds have already been described. They include processes such as molecular modelling, in which either the structure of the biological target molecule or a series of known compounds having known biological activity are utilised in planning new and better compounds. For various reasons, however, that process can be used only to a limited extent, especially when the structure of the biological target molecule is not known, or compounds having known activity are not yet

available.

Other processes (Agrafiotis, System and Method of Automatically Generating Chemical Compounds with Desired Properties, US 5 463 564, Oct. 31, 1995) require the complicated interpretation of structure-activity relationships, which in turn require the sequential, automated synthesis of purified compounds having a known structure. Such explicit structure-activity relationships have proved only of some use in the past, since although they can be used to predict to some extent an improvement in activity in a target molecule, they do not take account of other factors, such as, for example, oral bioavailability or toxicity, that exhibit a different dependency on the structure.

Furthermore, a process has been put forward which comprises the discovery and preparation of chemical compounds having desired properties without any knowledge of the structure of the synthesised compounds (S. Kauffman, J. Rebek, Random Chemistry for the Generation of New Compounds, WO 94/24314), but that process uses a large number of sequential synthesis steps in order to achieve molecular diversity. In that process the desired diversity is achieved only when the number of different chemical compounds in a synthesis and test sample reaches a high degree of complexity and exceeds a supercritical point. That process also does not overcome the problems of testing highly complex and unknown product mixtures which are known to lead to false positive and false negative results. Furthermore, the identification and isolation of chemical compounds that are present only at low concentration is technically complex; moreover the simple re-synthesis of a compound from such a supercritical mixture has not been described and ought to be difficult unless it involves biologically amplifiable compounds, such as peptides, DNA or RNA, that are of only

limited pharmaceutical importance.

A process which combines the search for and the preparation of chemical compounds with a complex target function that includes a plurality of desired properties or even a broad spectrum of properties (for the chemical target compound) has not yet been put forward.

On account of the problems described above there is a need for a process for the fast and efficient discovery and preparation of biologically active compounds which eliminates the disadvantages described.

There is also a need for a process that enables structurally new and natural-substance-analogous compounds to be prepared in a simple manner.

There is also a need for a process that enables those compounds to be tested in a simple manner, the results of the tests having as great as possible an influence on the further preparation of new, improved chemical substances.

A further objective of the present invention is a new means of access to new substance classes, such as polyketides, in only one or only a few step(s).

According to the invention there is provided a process for the fast and efficient discovery and preparation of biologically active compounds that achieves the objectives described above, especially objectives (a-h), and thus eliminates the shortcomings of known processes.

The process comprises the following steps:

(1) selection of M different starting materials suitable for multicomponent reactions (MCRs),

(2) reaction of each starting material with another of or with every possible combination of up to M-1 other starting materials selected according to (1),

(3) analysis of the products,

(4) evaluation of the products and selection of at least one product,

(5) determination of the starting materials that have led to the product(s) selected in (4), and

(6) provision of at least one variant of at least one of the starting materials that have been determined in (5),

(7) reaction of the starting materials provided in (6) if appropriate with the remaining starting materials determined in (5) in the context of an MCR,

(8) repetition of steps (4) to (7) until at least one product having the desired property or properties is found, and

(9) optionally isolation and characterisation of the product.

According to the invention, therefore, first of all a set of M different starting materials suitable for multicomponent reactions (MCRs) is selected. Those starting materials are then used to carry out all the MCRs that are possible with those starting materials, preferably at least three starting materials and at most all the starting materials being reacted with one another. Where there are five starting materials,

therefore, for example reactions of all three-component combinations, and all four-component combinations and one reaction of the five-component combination are possible.

The properties of the resulting products are then ascertained by analyses, assays etc., it not being absolutely necessary to analyse the products themselves or to clarify their structure. Such a step is of course possible, however, and also forms part of the present invention.

The properties of the products are then evaluated: if, for example, a product having an antibacterial action against *Pseudomonas aeruginosa* is found, then the product or products having the best such action is or are selected.

Those products or that product can then be compared if appropriate with the next best product or the next best products in order to form conclusions as to possible factors relevant to the action of the best product or products.

In order to establish which MCR has given rise to the optimum product, it can be compared with its sub- and supra-combinations:

If, for example, the product of a 5-component reaction has good properties, it is compared with the products of the 4-component reactions that are possible using those components and also with products prepared by 6-component reactions in which a further starting material is used in addition to the five starting materials used in the 5-component reaction. If, for example, the product of a 4-component reaction has similarly good properties to those of the product of the 5-component reaction, the obvious conclusion is that the fifth component does not contribute to the properties of the product and, for

example, acts only as a catalyst or does not participate in the reaction at all.

After the evaluation of the products, one or more products are selected.

The process according to the invention is described below with reference to a single selected product, it being clear that it is also possible to select a plurality of products and to carry out the process in parallel with a plurality of products.

In the next step, the starting materials that have led to the selected product having the best properties are determined. If, in the example given above, the product prepared from five starting materials has the best properties, then those five starting materials are taken as basis in the further steps of the process.

At least one of the starting materials, for example an amine component, is then varied or modified by, for example, replacing at least one substituent and/or introducing at least one (further) substituent. It will be understood that it is also possible for two, three or all starting materials to be chemically varied or modified.

The varied or modified starting materials are then reacted in the context of the same MCR as that which led to the optimum product in the preceding steps of the process. If, in the above example, two starting materials, for example, are modified in such a manner that there is a further variant of each of those two starting materials, but the other three starting materials are used further in unmodified form, then three new reactions of the same MCR type are possible, since in one reaction three original starting materials and two new

starting materials can be used and in the two other reactions four original starting materials and one of the modified starting materials can be used.

Then, in turn, the product or the products having the best properties is or are selected and if appropriate it is ascertained, on the basis of the variations or modifications, to which of those variations or modifications any improvement in the properties is to be attributed: if, for example, it is ascertained that the enlargement of a substituent on an amine component leads to an improvement in the properties, then the starting materials selected on repetition of the steps will be those in which at least one substituent on the amine component is further enlarged.

By means of the process according to the invention, therefore, for one specific set of starting materials in the first instance all the basic scaffolds that can be prepared by means of multicomponent reactions are prepared and the properties of the different scaffolds are compared with one another.

In the second step of the process, the substituents on the selected scaffold, for example, are subsequently altered until a product having an optimum action has been found.

One advantage of the process disclosed herein as compared with conventional methods of finding active ingredients is the fast and efficient discovery of chemical compounds that fulfil a desired target function. The target function may be, for example, a specific biological activity and/or a spectrum of other desired pharmacological and physico-chemical properties and is varied according to the molecule being sought etc.. The target function preferably involves therapeutic properties.

A further advantage of the process disclosed herein is that for the discovery of such compounds it is unnecessary to have any knowledge of the chemical structure of the compounds that have been prepared or are to be prepared or any knowledge of the chemical reactions taking place in the experiments. The setting-up of complex structure-activity relationships is therefore unnecessary.

A further advantage of the process disclosed herein is that the desired product can be prepared using a multicomponent reaction, whether it be by way of a chemical reaction that is already known or by way of a new chemical reaction found in the process according to the invention, and that the product is therefore accessible by means of very simple chemical reaction steps even when structurally very complicated compounds are involved.

A further advantage of the process disclosed herein is that the system of combining the different starting materials results in a large number of new and different basic chemical scaffolds, not only the substituents of those basic scaffolds but also the basic scaffolds themselves being varied and being tested for their suitability for finding new compounds having outstanding properties.

A further advantage is, in addition, that the efficiency of the process according to the invention in discovering and preparing biologically active chemical compounds can be measured in a simple manner.

Using the process according to the invention, therefore, the diversity, that is to say the information content of chemical reactions, chemical compounds or substance libraries, can, especially, be matched to the individual phases of the dis-

covery, optimisation and preparation of chemical compounds having the desired properties or can correspond thereto, especially since the advantages of combinatorial substance libraries are combined with those of random libraries.

According to the invention, therefore, there is disclosed a process that, in a plurality of cycles, enables new compounds having pharmacologically outstanding properties to be obtained quickly and efficiently by iterative selection of the starting materials, preparation of selected products by means of multi-component reactions and biological, pharmacological and/or physico-chemical testing thereof, especially testing thereof for their therapeutic potential.

The process can be used, for example, for the preparation of any products having desired properties, such as, for example, medicaments, veterinary products, vaccines, cosmetics, plant protection preparations etc. or additives thereto or ligands, catalysts, catalytic cofactors, detector molecules, polymers, peptides and adhesives.

The invention relates both to the process and to the products found by that process.

The process of the present invention will be described in detail below:

1) In a first step, therefore, a number **M** of different chemical starting materials is selected which are provided with functional groups customary in organic chemistry and suitable for multicomponent reactions (MRCs), such as Passerini or Ugi MCRs (J. March, Advanced Organic Chemistry, Wiley-Interscience, New York, **1984**), such as
-NC, -CO-, -CS-, -CN, -OCN, -NCO, -NO, $-NO_2$, $-ONO_2$, -CHO,

-COOR, -COSR, -CSSR, -COCOOR, -SCN, -NCS, -halo, $-N_3$, -NNNR, -OR, -SR, -OCOOR, -SCOOR, -NRCOOR´, -OCSOR, -SCSOR, -NRCSOR´, -OCSSR, -SCSSR, -NRCSSR´, -OCONR´R, -SCONR´R, -NRCONR´R``, -NRR´, -NRR´NR´´R´´´, -CNNRR´, -CNNRR´HX, -NRCONR´R´´, -NRCSNR´R´´, -RCOCR´R``, -RCSCR´R``, -COCRR-`halo, -RCNR`CR``, wherein R, R`, R`` and R''' may each independently of the other be H or alkyl, aryl, aralkyl, hetaryl or hetarylalkyl, "alkyl" preferably being C1-C10alkyl, "ar(yl)" preferably having up to 10, especially up to 2 or 3, preferably aromatic rings and "Het" preferably including N, O or S,
or epoxy groups or carbenes or unsaturated vinylogous variants (alkene, alkyne, aryl) of the above-mentioned functional groups, or corresponding mono-, di-, tri-, tetra-, penta- or hexa-carbonyl variants of the above-mentioned functional groups,
it being possible especially for two, three, four or more of the above-mentioned functional groups to be present simultaneously in one or more of those starting materials, especially in suitable combination.

Some of the functional groups can be provided with protecting groups customary in organic chemistry (T.W. Greene, **Protective Groups in Organic Synthesis**, Wiley-Interscience, New York, **1981**).

It is preferable to select those starting materials which are known to be good starting materials for multicomponent reactions, such as alpha-haloketones, esters, carboxylic acids, thiocarboxylic acids, aldehydes, amines, ketones, isonitriles, nitriles, alpha-keto acids, alpha-keto esters, and derivatives and alpha-beta unsaturated variants thereof, and combinations thereof,
special preference being given to corresponding mono-, di-, tri-, tetra-, penta- or hexa-carbonyl variants of the above-

mentioned functional groups.

According to the invention, those **M** starting materials are preferably encoded in a form accessible to an algorithm, the selected starting materials being assigned, either randomly or systematically, binary, decimal or alphanumeric codings.

Preferably a starting material type of a specific chemical class, such as, for example, aldehydes, is assigned a characteristic basic coding, such as, for example, "A", it being especially preferable for different starting materials that fall into that class, such as various special aldehydes, to be randomly assigned an additional coding, such as the numbers "1, 2, 3 …", giving rise to an alphanumeric overall coding A1 for benzaldehyde and A2 for acetaldehyde, or B1 for aniline and B2 for methylamine.

Chemical classes (substance classes, types of starting material) therefore denote, for example, aldehydes, amines, carboxylic acids, and they especially denote component groups of MCRs.

For **M** different starting materials there are thus obtained according to the invention **N** different codings. The set **N** is intended to indicate the set of (different) classes of starting material or chemical classes, it being possible according to the invention for a starting material to be assigned to several of those classes and encoded accordingly, such as, for example, beta-ketopropionic acid being assignable to the class of ketones, of carboxylic acids or of beta-keto acids. Special preference is given to a coding system in which each starting material is encoded only in one class.

It is especially possible, by suitable selection of starting

materials, to select or predetermine the product range, i.e. the nature and amount of the products provisionally to be obtained, at least within certain limits.

In the process according to the invention, preferably M is ≤ 40, more preferably M is ≤ 30, especially M is ≤ 20 and most preferably M is ≤ 12.

2) In a second process step, the starting materials are reacted simultaneously or sequentially in the context of an MCR (including an unknown MCR). In that reaction each starting material is reacted with every other starting material or preferably with every possible combination of from 2 up to M-1 other starting materials selected in the first process step.

An advantage of the process according to the invention lies in the fact that it is unnecessary to have any knowledge of the possible reactions into which those starting materials can enter.

According to the invention, in a second process step all the multicomponent combinations or, as the case may be, multi-component combinations MCC(K) selected in accordance with an algorithm of K different starting materials from a set N of different starting materials, which represents a sub-set of the set M of starting materials available, are reacted simultaneously or in a sequential order under conditions customary in organic chemistry, as customary, for example, for Passerini or Ugi MCR reactions with 4, 5, 6, 7, 8, 9 or 10 components. For that purpose the starting materials or the selected starting materials can be combined in one or more solvents, such as methanol, tetrahydrofuran, dioxane, dimethyl sulfoxide, water or mixtures thereof, if necessary with the exclusion of air or under a nitrogen, oxygen, hydrogen or argon atmosphere, in a

temperature range of between -60°C and 150°C. In addition, it is possible to use auxiliaries or catalysts, such as, for example, Lewis acids, such as boron trifluoride etherate, zinc chloride, ytterbium triflate, iron chloride, other acids, such as, for example, hydrochloric acid, paratoluenesulfonic acid, acetic acid, or bases, such as, for example, potassium carbonate, triethylamine, caesium carbonate, or water-removing agents, such as molecular sieves or orthoesters.

In the first cycle of the process, from the set **M** there are preferably selected those **N** which belong to different substance classes,
the total number of all experiments **E** being the sum of **K** = 1 to **N** over all **K** from **N** according to equation (1)

$$E = \sum N!/((N-K)!*K!) \qquad \text{equation (1),}$$

where **K** is accordingly the number of different starting materials used for a reaction, and **N** is the highest possible number of different starting materials used in a reaction,
it being especially preferable in the first cycle of the process for all combinations from **K** = 1 to **K** = **N** to be selected.

Each of those experiments can be carried out physically separately and in a reproducible manner, for example in different reaction vessels, and especially the allocation of the different combinations with their codings to the positions of the reaction vessels can be stored in a computer in a form accessible to an algorithm.

At least some of the resulting reaction products can in a subsequent step be further chemically modified, worked-up or prepared for step (3) in a suitable manner.

Such a chemical modification may be, for example, the removal of chemical protecting groups, for example by trifluoroacetic acid, or the hydrogenation of the products by means of hydrogen, optionally with the addition of a hydrogenation catalyst, such as, for example, palladium on carbon, platinum oxide, palladium acetate, or by oxidation of the products with oxygen or some other oxidising agent, such as, for example, bromine, hydrogen peroxide, tert-butyl peroxide or a suitable metal salt, such as, for example, cobalt chloride, or a suitable metal complex, such as, for example, iron hexacyano-ferrate or chromium tetraphenylporphyrinate, or by irradiation with light of wavelength 200-600 nm. Furthermore, the reaction products can be treated with one or more enzymes, such as, for example, oxidoreductases, ligases, peptidases, lipases or isomerases.

The working-up of the products can be carried out in a manner known *per se*, such as by chromatography, for example over silica gel or RP-18 silica gel, or solid phase extraction or the removal of unreacted starting materials by binding to a suitable solid carrier, such as, for example, ion exchanger resins or chemically modified solid phase resins, or alternatively the expected products can be purified by selective binding to such a solid carrier, followed by washing and detachment from that carrier.

By subsequent dissolution in a suitable solvent, such as, for example, water or DMSO, a test solution can be prepared.

The reaction conditions, modifications, working-up procedures or procedures in preparation for testing that are used can likewise be encoded in a form suitable for an algorithm, for example in binary, decimal or alphanumeric form. A reaction

product can accordingly be encoded, for example, either as a combination of the coding of the starting materials used or preferably as a combination of the coding of the starting materials used and the coding for the reaction conditions, modifications, working-up procedures or procedures in preparation for testing that are used, it being especially preferable that both the starting materials, the reaction conditions, modifications, working-up procedures or procedures in preparation for testing that are used and the reaction vessels be encoded.

Such a coding will be referred to below simply as the genome of the reaction product.

For the purposes of the process described herein it is not necessary to know which reactions may take place or do take place in the individual reaction vessel. According to the invention, however, in all reaction vessels a maximum of E different reaction types and accordingly E different chemical substances each having different basic scaffolds may be formed when all the starting materials are selected from different substance classes, as is preferred in the first cycle of the process. The genome of a reaction product especially preferably encodes not what is contained in a reaction vessel, but by means of which starting materials and using which process steps the reaction product has been formed.

3) In a third process step, for example, test solutions of the products from the second process step are investigated, for example, in a biological and/or pharmacological and/or physico-chemical test for their biological activity, effectiveness, side effects or selectivity and/or in another test procedure the physico-chemical properties of those products are investigated.

Those biological, pharmacological and physico-chemical test procedures are known *per se* to a person skilled in the relevant art.

In those procedures, it is preferable especially to investigate and ascertain the dependency of the measurement results upon the concentration of the starting materials used in process step two. Especially a concentration range of from 0.5 to 0.000001 mol/l, more especially a concentration range of from 100 to 0.01 mol/l, is investigated.

The test for ascertaining the biological or pharmacological activity, effectiveness, side effects or selectivity is preferably carried out with isolated proteins, receptors, enzymes, or mixtures thereof, cells, cell lysates, complex cell systems, with organs or parts thereof or a plurality of organs or alternatively with whole organisms or membranes and as appropriate using adjuvants, substrates or detection aids necessary for the test.

The test procedures for the physico-chemical properties of the products may include, for example, the measurement of the lipophilicity by means of the octanol-water distribution coefficient, the solubility in water, the non-specific protein binding to, for example, bovine serum albumin, the binding to the proteins of human serum plasma or the chemical stability in Krebs buffer.

The test results obtained are preferably correlated with the genomes of the reaction products, especially in a form accessible to the algorithm, for example in a computer data file or a computer data bank.

According to the invention, for the purposes of the process it is not necessary to have any knowledge of the content of the individual reaction products, such as, for example, including knowledge of the chemical reaction taking place or the new chemical compounds present and the structure thereof, since the systematic nature of the selection allows a systematic interpretation of the test results. It may even be possible that no reaction at all has taken place in one or more of the parallel reaction vessels, without it being a disadvantage for the process according to the invention.

If, for example, all combinations of from **K**=1 to **K=N** have been selected, according to the invention all starting materials (K=1) are tested for their biological action. All reaction products containing those starting materials but having a better action than the latter ought to contain a new chemical compound having a better action. The same is analogously true also of all combinations K=3 and all two-component combinations. A three-component combination that exhibits an action better than that of the two-component combinations it contains or better than that of the corresponding starting materials, ought to contain a new, effective chemical compound from a three-component reaction. All **K**>2 reaction products can likewise be analysed in the manner indicated.

According to the invention, the list of the genomes and the associated test results contain all the information necessary for further optimisation.

The process according to the invention can implicitly use a statistical analysis of the reactions and working steps carried out, the system used in the selection of the starting materials **M** making it unnecessary to have exact and explicit knowledge of the chemical reactions that have taken place and the structures

of the resulting compounds. For example, it is possible that a reaction that is known and desirable *per se* does not take place under the reaction conditions employed, but a different, previously unknown reaction yields a new chemical compound having desirable properties, such as, for example, oral bioavailability. The corresponding genome of that reaction product and the associated test results therefore implicitly also includes the process, as well as the yield and structure of the chemical compound from that new multicomponent reaction. As a result it is possible to use that reaction, even without having explicit knowledge of it, with the algorithm according to the invention.

4) In a fourth process step, the test results measured for the products prepared are used for evaluating the products, which are preferably encoded, and, for example, for sorting them in accordance with a predetermined target function, and selecting at least one product,
it being possible for that target function to be any combination of desired properties for the target compound being sought and for the sorting criterion to be derived from the extent to which the individual products fulfil that target function. Preferably the products are evaluated according to their concentration dependency.

The products can especially either be sorted by ranking or divided into various evaluation categories.

The target function can be any function that is construed for the target compound sought from the combination of the desired properties in the test systems used. It is the evaluation criterion for the sorting or categorisation of the genomes according to the manner in which the respective individual products fulfil that target function.

Preferably the biological activity, physico-chemical properties as well as further biologically relevant test results form that target function.

It is especially preferable that the concentration dependency of the test results be ascertained and included in the target function, that is to say that those properties are included in the target function with a different and concentration-dependent weighting.

The target function is especially preferably a linear combination or a polynome of those properties with "fuzzy" logic weightings, the "fuzzy" logic weightings of individual properties being especially dependent upon the extent to which other properties are fulfilled and upon the number of cycles already completed.

According to the invention, such a target function can accordingly assume the form of a program which differently evaluates a genome in dependence upon various properties and conditions with logical and conditional links between different evaluation functions. For example, according to the invention a high rating may be attached to those genomes which initially have, for example, a high oral bioavailability, or after several cycles have a plurality of those desirable properties. Equally, according to the invention compounds having some desirable properties may, however, also have properties that have been designated as undesirable, such as, for example, a measured logD value for the lipophilicity of more than 5 may be given a negative rating, in which case the desirable properties are no longer considered.

5) In a fifth process step, the starting materials that have

led to the product(s) evaluated and selected in the fourth step are determined.

In this step it is unnecessary to analyse the starting materials themselves or to clarify their structure: rather, it is sufficient to identify the starting materials by reference to any coding used, since each starting material can be assigned a specific code.

6) In a sixth process step, according to the invention a new set of starting materials is chosen on the basis of the found results, for example using an algorithm. Furthermore, a list of fresh experiments to be carried out is drawn up, and the starting materials are combined and reacted in correspondingly selected multicomponent combinations, but an experiment that has already been performed is preferably not proposed again.

For that purpose there is provided a variant or modification of at least one, preferably two, of the starting materials determined in process step (5) in the sense that in that starting material at least one substituent is exchanged and/or at least one (additional) substituent is introduced and/or an existing substituent is replaced by a hydrogen atom.

Preferably, for each cycle more than one starting material and/or more than one reaction parameter, such as the concentration of a starting material or the reaction temperature etc., is varied.

On the basis of combinatorial optimisation procedures known *per se*, see Cook, W.J.; Cunningham, W.H.; Pulley-blank, W.R. and Schrijver, A. Combinatorial Optimization, Wiley 1997; Philip M. Dean and Richard A. Lewis (Ed.) Molecular Diversity in Drug Design, Kluwer Academic Publishers, 1999, it is possible to

assign altered product properties to the varied starting materials and/or reaction parameters.

Preferably the genomes of the preceding cycle that are evaluated as being the best are used for the generation of the new genomes.

As algorithm it is possible to use, for example, a combinatorial optimisation procedure, such as a genetic algorithm or a pattern recognition process, such as, for example, a neuronal network or a combination of a genetic algorithm with a neuronal network,
a genetic algorithm or a pattern recognition process, such as, for example, a neuronal network or a combination of a genetic algorithm with a neuronal network, preferably implicitly or explicitly correlating the occurrence of desired properties with the constituents of the product genome of the preceding generation.

It is preferable that those constituents of the genomes of the tested products which with greater probability correlate explicitly or implicitly with the desired properties be used with greater probability for the generation of the new genomes, those genomes the products of which have not received good ratings preferably not being used for the generation of new genomes, and
preferably individual constituents of the new genomes being selected randomly from the number of possible codings by means of a random generator.

Preferably, individual constituents of the new genomes are randomly removed from or added to the genome by means of a random generator,

the assignment of probability of a random selection of such a building block preferably being dependent upon the type of that building block,
the genomes especially being divided randomly into one or more groups, so-called populations.

Especially preferably, the genomes of a group are used only for the generation of new genomes of a new group of genomes and thus each of those populations will create a new population, preferably it being possible after any desired number of cycles for all populations of genomes to be divided up into a new number of populations having the same number or a different number of genomes.

Especially preferably, that new division is carried out when in a population a product has especially desirable properties.

Constituents of the genomes as defined according to the invention include the different codings of the starting materials, of the reaction conditions, modifications, working-up procedures or procedures in preparation for testing and also the reaction vessels.

According to the invention, that process step constitutes a transfer of the natural evolution of biopolymers, such as DNA, RNA or peptides, to the chemistry of multicomponent reactions in conjunction with the properties of the chemical compounds thereby created. According to the invention, that step enables the number of possible reactions to be appreciably increased. Since, by means of the algorithm according to the invention, building blocks of the genome can be deleted or added as desired, multiple uses of a building block, such as, for example, a starting material, a catalyst etc., for example, are possible.

The special nature of the algorithm lies in the fact firstly that preference is given to those genomes, that is to say those combinations of starting materials, reaction conditions, modifications, working-up procedures or procedures in preparation for testing, the products of which also have desired properties, and secondly that although those compounds are unknown the best reaction conditions for their actual preparation are also implicitly determined. Although the reaction leading to that product need not be known, that reaction is optimised because, for example, a higher yield of the product obtained by that reaction is manifested in an improvement in the desired properties. As a result of this characteristic of the process according to the invention there is thus obtained, in addition to a product having the desired properties, at the same time also its optimum method of preparation by means of a multicomponent reaction.

As desired, the starting materials and/or reactions/reaction conditions can each be varied individually or some or all of them can be varied together.

7) In a seventh process step, the starting materials provided in the sixth process step are reacted if appropriate together with the remaining starting materials determined in process step (5):

If, for example, only one starting material has been varied in process step (6), then that starting material is preferably reacted with the remaining starting materials determined in process step (5) with the exception of the starting material that was varied in process step (6).

If, in process step (5), the starting materials E_1, E_2, E_3, E_4

and E_5, for example, were determined, and if E_2 was varied to form E_2' in process step (6), then in process step (7) E_2' is reacted with E_1, E_3, E_4 and E_5. Preferably, in the reaction only one molecule is used per starting material type, that is to say only one amine, one isocyanide, one carboxylic acid compound.

8) In an eighth process step, process steps four to seven are repeated until a reaction product fulfilling the criteria of the target function is found,
frequently up to 50 cycles, especially up to 30 cycles, being required to find such a product.

The probability of discovering such a product can be estimated after as few as 2 to 6 cycles, so that a route showing little prospect of success can be terminated at an early stage.

Preferably there are used for that estimation the difference between the average extent to which the products of a genome population from a cycle **x** fulfil the target criteria and the average extent to which the products of a genome population from a later cycle x+i fulfil the target criteria, where i is a whole natural number.

That difference can be used to select a new number of starting materials and to begin the iterative process according to the invention afresh, especially when that difference is small.

By means of process steps one to eight, various problems in the discovery and optimisation of new chemical compounds are solved simultaneously in a new and surprising manner. By combining starting materials from different substance classes under different reaction conditions, new multicomponent reactions are investigated for their suitability for the preparation of new

chemical compounds, preference being given to those multicomponent reactions which yield products having desired properties. Furthermore, according to the invention those products are varied by the possibility of using starting materials from the same substance class that are necessary for those multicomponent reactions, and tested for their properties and optimised. Moreover, according to the invention those new multicomponent reactions are themselves optimised when the reaction conditions are constituents of the corresponding genomes.

9) In a ninth process step, the chemical compounds contained in the reaction product that has exhibited the desired properties in the tests are purified in a manner known *per se*, such as, for example, by chromatography or crystallisation, and the structure thereof is determined using known methods, such as mass spectroscopy or NMR spectroscopy.

The novel process will be described by way of the example of the discovery and preparation of a very great variety of non-natural antibiotic, immunosuppressive, antineoplastic or anthelmintic polyketoidal compounds having desired properties in order to clarify its advantages over existing processes.

Polyketides are a structurally highly diverse family of natural substances which are synthesised in nature by a common biosynthesis route. The family of polyketides has provided an extraordinarily large number of substances having interesting biological activities. For example, many examples of polyketides are cancer drugs, antibiotics, anthelmintics, immunosuppressives or the like. Prominent commercially available examples are the tetracycline antibiotics, FK 506 and rapamycin, adriamycin and epothilon, or monensin (Figure 1).

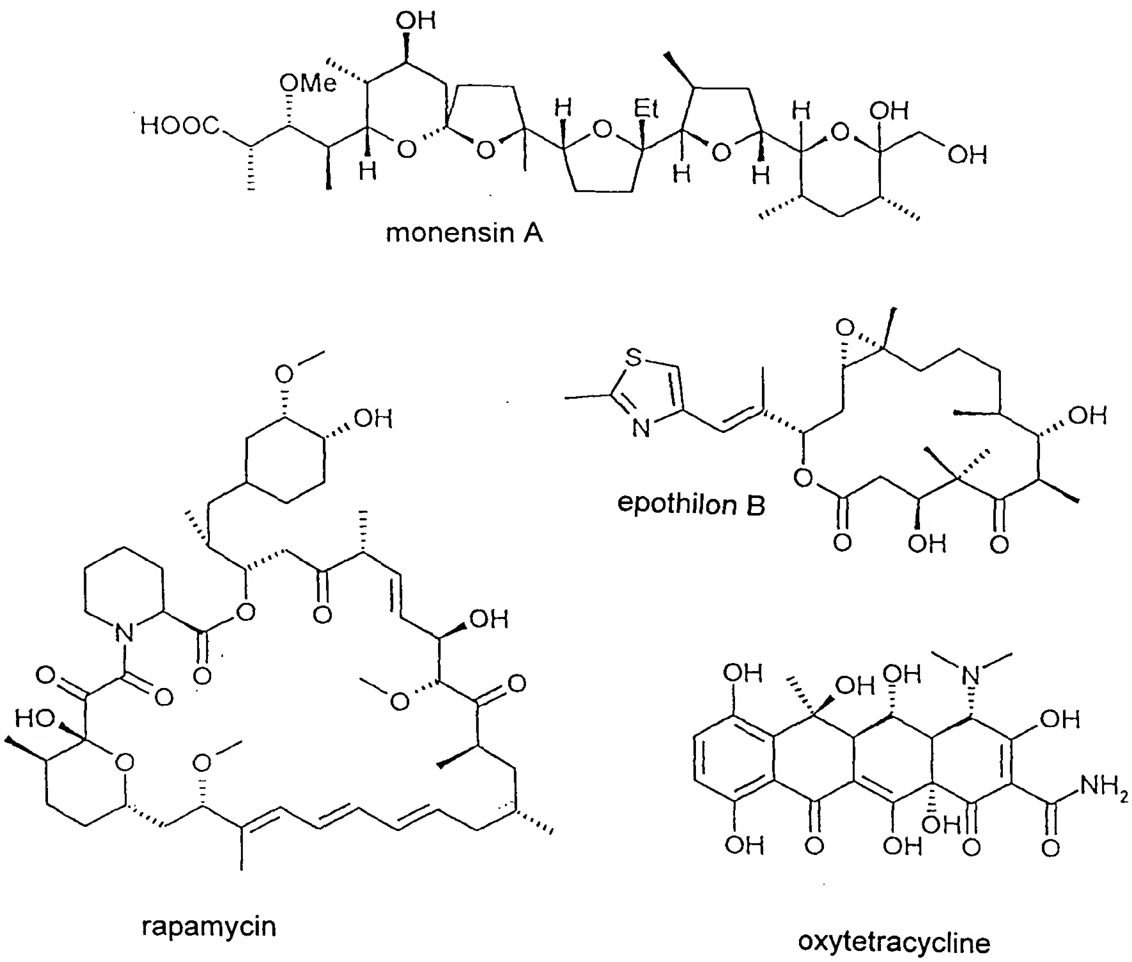

Figure 1: Various structures of polyketides.

Polyketides are formed by almost all classes of organism, but especially by mycelium-forming bacteria of the Actinomyces class.

In nature, polyketides are synthesised *via* the so-called polyketide route, in which putative polyketides are assumed to be intermediates of biosynthesis (Figure 2).

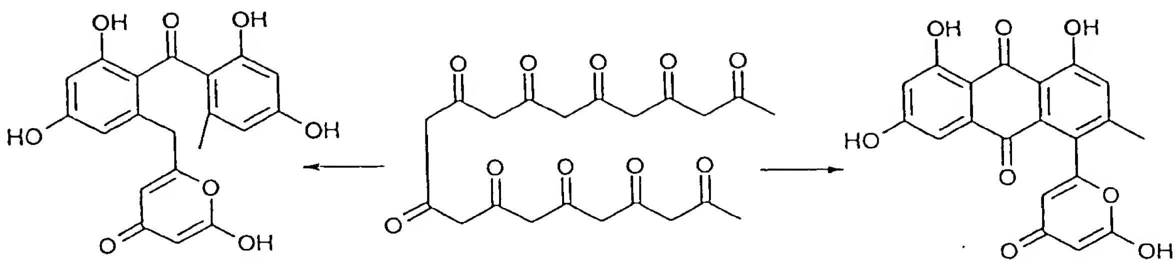

Figure 2: A polyketide precursor which, according to the nature of the cyclisation, results in different products.

Polyketide synthases (PKSs) are multifunctional enzyme complexes that are related to fatty acid synthases. The structural variety of polyketides comes about as a result of repetitive synthesis *via* decarboxylating Claisen condensation between different thioesters (mainly acetyl, proprionyl, butyryl, malonyl, methylmalonyl) to form polyketides and modifications thereof, such as, for example, reduction to alcohols, dehydration etc.. Each product of the polyketide synthesis route comes about as a result of a characteristic number of cycles, the product being split off at the end of the synthesis, frequently with cyclisation by the PKS.

Accordingly the diversity of this group of substances is brought about by the starter thioester, the reductive cycles and the number of decarboxylating condensation cycles.

A distinction is drawn between two classes of PKSs. The first class of type I is capable of synthesising complex macrolides, such as, for example, erythromycin. The second class of type II is capable of synthesising aromatic products.

Recently some working groups have been successful in synthesising new polyketides not previously found in nature by

means of genetically manipulated PKSs (Khosla, Leadley, Katz, Chem. Rev. **97**, 97,7).

The chemical synthesis of many polyketides has also been and is being explored by many working groups (Harris, T.M., Harris, C.M., *Pure & Appl. Chem.*, **1986**, *58*, 283 - 294. Alternatively Nicolaou, K.C.; Vourloumis, D.; Li, T.;Pastor, J.; Winssinger, N.; He, Y.; Ninkovic, S.; Sarabia, F.; Vallberg, H.; Roschangar, F.; King, N.P.; Finlay, M.R.V.; Giannakakou, P.; Verdier-Pinard, P.; Hamel, E. *Angew. Chem., Int. Ed. Engl.* **1997**, *36*, 2097). In all cases the syntheses have a large number of steps, are time-consuming and have little variability and have low total yields. Furthermore, combined biosynthetic-synthetic routes are followed, in which fermented polyketides are subsequently chemically modified. None of the chemical set-ups enables polyketides to be synthesised in sufficient variety and in a sufficiently small number of steps.

All the chemical routes followed hitherto are therefore tedious, difficult, expensive and unsuitable for fast, efficient and economical discovery of new polyketoidal active ingredients.

Examples:

The new process is described *by way of example* for the preparation of a very great variety of non-natural antibiotic, immunosuppressive, antineoplastic or anthelmintic polyketides.

Example 1: Preparation of a substance library of different multicomponent reactions having antibacterial action.

10 starting materials **1-10** (see Figure 3) having different functional groups are selected: benzaldehyde **1**, aniline **2**, 3-

phenyl-3-keto-propionic acid ethyl ester **3**, 2,4-diketovaleric acid ethyl ester **4**, 3-keto-glutaric acid dimethyl ester **5**, 2-keto-propionaldehyde **6**, 3-methyl-2,4-diketopentane **7**, 3,5-di-keto-5-phenyl-valeric acid **8**, 2,4-diketo-phenyl-butyric acid **9** and diphenylmethaneisonitrile **10**.

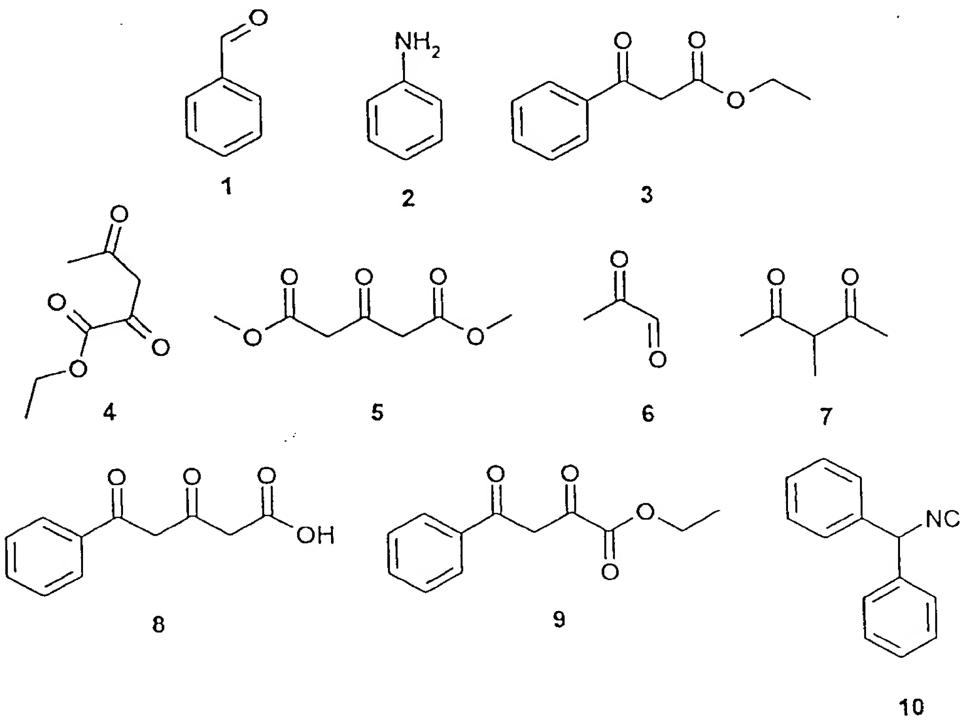

Figure 3: The selected starting materials for a combinatorial library of 1023 different multicomponent reactions.

The systematic variation of the starting materials from **K**=2 to **K**=10 yields, in accordance with equation (1), 1013 different possible ways of combining the starting materials:

Schedule 1:

Number of reactants	Number of combinations
2	45
3	120
4	210
5	252
6	210
7	120
8	45
9	10
10	1

$\sum$ = 1013 reactions

The selection of those starting materials allows reactions known *per se*, such as, for example, the Ugi 4-CR and 3-CR reaction, and various aldol and Claisen reactions, and also cyclisation reactions in accordance with Figure 2.

The 10 starting materials were prepared for the combination of the individual reactions in the form of 0.05M solutions in ethanol. The 1013 different possible combinations were carried out under four different reaction conditions (Set A, B, C, D). For the 4*1013 different reaction batches 20 µl of the respective starting material solution were dispensed in each case. Reaction set A of 1013 parallel batches was carried out without further additives. In the case of reaction set B, 10 µl of a 0.2M solution of p-toluenesulfonic acid in EtOH were additionally added in each case. In the case of reaction step C, 10 µl of a 0.2M solution of triethylamine in EtOH were additionally added. In the case of reaction set D, 10 µl of a 0.2M solution of potassium carbonate in a 2:1=EtOH:water mixture were additionally added.

When the addition was complete, the in total 4052 reactions were left to stand in a closed container at room temperature

for 24 hours. The solvent was then evaporated off at room temperature. The crude products were each diluted with 250 µl of DMSO, and a 10 µl portion of each resulting solution was diluted with 140 µl of water. The resulting solutions were tested for their inhibitory action with respect to gram-positive and gram-negative bacteria and yeast strains. The test results provided information about those reaction batches, or reaction types, which are of interest for further optimisation. Table 1 lists the test results using the example of the action with respect to Pseudomonas aeruginosa ATCC 9027 and Staphylococcus aureus ATCC 6538. The test organisms were cultured overnight in CASO bouillon (bacteria) at 35°C or Sabourad bouillon (yeasts) at 22°C. The suspension of the organisms was centrifuged off, the pellet was resuspended in fresh medium and incubated for a further 2 hours. The pellet was then resuspended in 0.9 % NaCl solution and the cell count was adjusted with reference to the standard curves to about 10^8 CFU/ml (bacteria) and 10^7 CFU/ml (yeasts).

The suspensions so obtained were then diluted to about 10^6 CFU/ml in CASO bouillon (bacteria) or Sabourad bouillon. 15 µl of the solution of the reaction products were inoculated with 100 µl of the resulting organism solutions. Immediately after the inoculation, and 7 and 22 hours incubation of the plates, the latter were measured in a plate reader (Bio-tek EL 311 Autoreader) at 550 nm.

The 1013 different combinations of starting materials 1-10 are listed in Schedule 1, and the inhibitory activity of the best starting material combinations of reaction set A after one cycle of the process according to the invention with respect to Pseudomonas aeruginosa are given in Table 1 and with respect to Staphylococcus aureus are given in Table 2.

From the best combinations, those to be used in the next cycle can then be selected, for example, in accordance with one of the algorithms described above.

In an analogous manner, the results of reaction sets A, B, C and D are compared with one another and the best reaction variants are considered in a corresponding manner in the next cycle of the process.

In summary, a process for the algorithmic discovery and preparation of biologically active chemical compounds is described. The process consists of the (1) production of an algorithmic library of different multicomponent reactions, starting from a library of suitable and diverse types of chemical starting materials, the (2) biological testing of that library, the (3) identification of suitable multicomponent reactions from that range of possible reactions, the (4) selection of a plurality of chemical starting materials of the types required for the identified and suitable multicomponent reactions, the (5) discovery of optimum combinations from the so constructed chemical range of those suitable multicomponent reactions by the (6) algorithmic preparation and biological testing of compounds from that library. The process is described by way of the example of the discovery of new antibiotically effective polyketoid-type compounds.

Table 1: The inhibitory activity with respect to Pseudomonas aeruginosa of reaction set A of 1013 different reactions of starting materials 1-10.

Table 2: The inhibitory activity with respect to Staphylococcus aureus of reaction set A of 1013 different reactions of starting materials 1-10.

Table 3: The ranking of inhibitory activity with respect to Pseudomonas aeruginosa of the best starting material combinations of reaction set A after one cycle of the process according to the invention.

Table 4: The ranking of inhibitory activity with respect to Staphylococcus aureus of the best starting material combinations of reaction set A after one cycle of the process according to the invention.